

①⑨ RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
PARIS

①⑪ N° de publication :
(à n'utiliser que pour les
commandes de reproduction)

2 799 914

②① N° d'enregistrement national : 99 12942

⑤① Int Cl⁷ : H 04 M 3/56, H 04 Q 7/20

⑫

DEMANDE DE BREVET D'INVENTION

A1

②② Date de dépôt : 14.10.99.

③⑦ Priorité :

④③ Date de mise à la disposition du public de la
demande : 20.04.01 Bulletin 01/16.

⑤⑥ Liste des documents cités dans le rapport de
recherche préliminaire : *Se reporter à la fin du
présent fascicule*

⑥⑦ Références à d'autres documents nationaux
apparentés :

⑦① Demandeur(s) : FRANCE TELECOM Société ano-
nyme — FR.

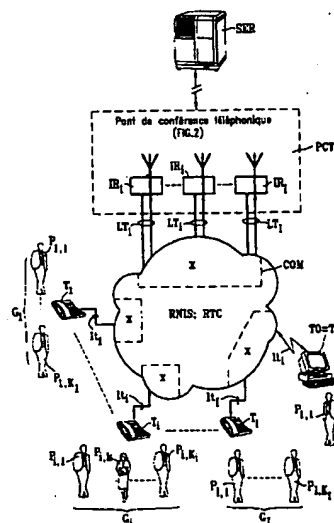
⑦② Inventeur(s) : LANIEPCE SYLVIE et CHARLET DEL-
PHINE.

⑦③ Titulaire(s) :

⑦④ Mandataire(s) : MARTINET ET LAPOUX.

⑤④ IDENTIFICATION D'INTERVENANT DANS UNE TELEREUNION.

⑤⑦ Un intervenant parmi des participants répartis en des
groupes (G_1 à G_i) à proximité de terminaux téléphoniques
(T_1 à T_i) est identifié par la voix lors d'une téléconférence gérée
par un dispositif de conférence (PCT) relié aux terminaux.
Après identification de la voix de l'intervenant en réception
par le dispositif de conférence, un identifiant de l'intervenant
identifié est diffusé depuis le dispositif de conférence vers
les terminaux et restitué dans les terminaux. L'intervenant
est présenté aux autres participants, automatiquement ou à
la demande, par exemple par un message vocal ou visuel
comportant son nom. Le dispositif de conférence peut être
inclus dans l'un des terminaux; par exemple, l'identifiant
identifie un intervenant autre que le propriétaire d'un poste
téléphonique ou radiotéléphonique en communication.



FR 2 799 914 - A1



Identification d'intervenant dans une téléréunion

La présente invention concerne les téléréunions vocales, particulièrement des conférences téléphoniques, organisées entre au moins deux groupes de participants éloignés.

Elle concerne plus particulièrement la reconnaissance de la voix d'un intervenant au cours d'une téléréunion indépendamment du contenu de son intervention afin que les autres participants à la téléréunion l'identifient. En d'autres termes, il s'agit de répondre à la question au cours d'une téléréunion "Quelle est la personne qui a parlé ou qui parle parmi tous les participants ?".

L'identification vocale, appelée également indexation de locuteur, consistant à reconnaître l'identité d'un locuteur dans un signal de voies de parole mélangées de plusieurs participants est déjà connue par l'article intitulé "REAL TIME SPEAKER INDEXING BASED ON SUBSPACE METHOD - APPLICATION TO TV NEWS ARTICLES AND DEBATE" de M. Nishida et Y. Ariki, Labs - 5^{ème} ICSLP, Sydney, Australie, décembre 1998.

Selon cet article, pour chaque locuteur est créé préalablement un modèle vocal de référence. Puis pour reconnaître la voix d'un locuteur, par exemple la voix d'un participant dans un débat télévisé, le signal des voies mélangées relatives aux voix des participants est découpé périodiquement en segments vocaux. Les distances entre chaque segment vocal et les modèles vocaux sont comparées pour ne sélectionner que la plus petite distance sous réserve qu'elle soit supérieure à un seuil. Le locuteur

correspondant à la plus petite distance est ainsi identifié comme l'intervenant pour le segment vocal.

Cependant, dans la configuration selon l'article précité, un intervenant est reconnaissable que dans
5 des voix mélangées ne provenant que d'un faible nombre de personnes qui sont physiquement proches les unes des autres. Cette identification d'intervenant est d'autant moins performante que le nombre de participants est élevé.

10

La présente invention vise à identifier un intervenant parmi plusieurs groupes de participants qui sont éloignés dans le cadre d'une téléconférence, c'est-à-dire des groupes situés à proximité de
15 terminaux éloignés, afin que chaque groupe puisse répondre plus précisément à la question "qui parle?". L'invention apporte ainsi une solution à la présentation des intervenants dans une téléconférence où le nombre de participants peut être élevé, en
20 s'affranchissant de toute confusion sur l'identité de ceux-ci.

A cette fin, un procédé pour identifier un intervenant parmi des participants répartis en des
25 groupes à proximité de terminaux téléphoniques lors d'une téléconférence gérée par l'intermédiaire d'un dispositif de conférence relié aux terminaux à travers des liaisons de transmission est caractérisé en ce qu'il comprend, après identification de la voix
30 de l'intervenant en réception dans les liaisons par le dispositif de conférence les étapes de :

diffuser un identifiant de l'intervenant identifié dans les liaisons depuis le dispositif de conférence vers les terminaux, et

restituer l'identifiant de l'intervenant dans les terminaux.

Grâce à l'invention, l'identité des intervenants au cours de la télé Réunion est indiquée à tous les participants au fur et à mesure des prises de parole.

L'identifiant de l'intervenant comprend au moins le nom du participant et peut être de préférence complété par au moins une désignation du groupe auquel il appartient, ou le nom de la société de l'intervenant, ou de toute autre information utile à l'identification de l'intervenant pour les participants. L'identifiant d'un participant peut être obtenu par reconnaissance de parole dans le dispositif de conférence après que celui-ci ait analysé une phrase d'apprentissage contenant l'identifiant prononcé par le participant et ainsi ait constitué un modèle vocal de l'identifiant vérifié par le participant. L'identifiant peut être également saisi au clavier d'un terminal pour sa reconnaissance vocale ultérieure.

En cas d'échanges conversationnels rapides pouvant engendrer des difficultés d'identification vocale, l'invention prévoit que l'identifiant ne soit diffusé que vers des terminaux en réponse respectivement à des demandes de ceux-ci au dispositif de conférence.

Selon une première variante, l'identifiant de l'intervenant est diffusé par le dispositif de conférence sous forme de message vocal en coupure ou en superposition de signaux vocaux de participant dans les liaisons et est restitué acoustiquement par les terminaux. Selon une deuxième variante, l'identifiant de l'intervenant est diffusé par le dispositif de conférence sous forme de message de signalisation dans les liaisons et est restitué

visuellement dans les terminaux par afficheurs ou écrans de ceux-ci.

Selon certaines applications, le dispositif de conférence, c'est-à-dire les fonctions principales de celui-ci, peut constituer un pont de conférence
5 amélioré, ou peut être inclus dans l'un des terminaux, ou dans un commutateur téléphonique privé desservant les terminaux.

Par exemple, lors d'une communication téléphonique entre deux postes téléphoniques ou
10 radiotéléphoniques, un participant devant l'un des postes équipé du dispositif de conférence a connaissance de l'identité de l'intervenant en tant que l'un d'au moins deux participants devant l'autre
15 poste.

Selon une autre caractéristique de l'invention, qui est de préférence à mettre en oeuvre pour un nombre de groupes de participants élevé,
20 l'identification de l'intervenant dans le dispositif de conférence comprend les étapes de :

- rechercher en réception la plus active des liaisons,
- identifier le groupe correspondant à la plus
25 active liaison en réception,
- déterminer la plus grande similitude entre un segment vocal dans la plus active liaison et des modèles vocaux des participants appartenant au groupe identifié, et
- 30 - identifier le participant correspondant à la plus grande similitude déterminée, en tant qu'intervenant.

Grâce à la caractéristique ci-dessus, la recherche de la voix d'un intervenant est effectuée
35 dans l'ensemble des voix des participants appartenant

au groupe identifié, et non parmi les voix de tous les participants disséminés autour des terminaux, ce qui rend plus fiable l'identification d'intervenant.

5 L'étape de rechercher en réception la plus active des liaisons peut consister à rechercher la plus grande des puissances moyennes de segments vocaux dans les voies de réception des lignes de transmission à condition qu'elle soit supérieure à un seuil prédéterminé.

10 Lorsque la plus grande similitude déterminée est inférieure à un seuil prédéterminé, le participant correspondant à la plus grande similitude déterminée peut être considéré comme non identifié, un message d'intervenant inconnu étant alors diffusé vers les
15 terminaux.

Les modèles vocaux des participants sont créés et mis à jour dans le dispositif de conférence au plus tard au début de la télé Réunion, et peuvent résulter de l'analyse vocale d'un genre de phrase
20 excédant une durée minimale prédéterminée, de préférence contenant au moins une partie d'identifiant de participant.

Le procédé d'identification peut comprendre une étape de déterminer la similitude entre le segment
25 vocal dans la plus active liaison en réception et un modèle vocal de prise de parole multiple, de préférence spécifique au nombre de participants dans le groupe identifié, afin que le dispositif de conférence diffuse un message de prise de parole
30 multiple vers les terminaux lorsque ladite similitude déterminée est supérieure à un seuil prédéterminé.

Le procédé d'identification peut encore comprendre une étape de déterminer la similitude entre le segment vocal dans la plus active liaison en
35 réception et le modèle vocal d'un dernier intervenant

afin que le dispositif de conférence ne diffuse aucun identifiant, ou ne diffuse l'identifiant du dernier intervenant que périodiquement vers les terminaux ou que vers des terminaux respectivement en réponse à
5 des demandes de ceux-ci au dispositif de conférence, lorsque ladite similitude est supérieure à un seuil prédéterminé.

L'invention concerne également un dispositif de conférence qui comprend un moyen pour détecter
10 périodiquement la plus active liaison en réception afin d'identifier le groupe correspondant à la plus active liaison en réception, un moyen pour comparer périodiquement un segment vocal dans la plus active liaison en réception avec des modèles vocaux
15 mémorisés des participants appartenant au groupe identifié afin d'identifier le participant, en tant qu'intervenant, correspondant à la plus grande similitude entre ledit segment vocal et les modèles vocaux, un moyen pour établir un identifiant de
20 l'intervenant identifié, et un moyen pour diffuser l'identifiant d'intervenant dans des liaisons vers des terminaux.

D'autres caractéristiques et avantages de la
25 présente invention apparaîtront plus clairement à la lecture de la description suivante de plusieurs réalisations préférées de l'invention en référence aux dessins annexés correspondants dans lesquels :

- la figure 1 est un bloc-diagramme schématique
30 d'un système de téléréunion de type conférence téléphonique selon l'invention ;

- la figure 2 est un bloc-diagramme schématique d'un pont de conférence inclus dans le système de la figure 1 ;

- la figure 3 est un algorithme d'identification d'intervenant selon l'invention, mis en œuvre dans le système de la figure 1 ;

5 - la figure 4 est un algorithme de création de modèle vocal de participant inclus dans l'algorithme de la figure 3 ; et

- la figure 5 est un algorithme de recherche et d'identification d'intervenant au cours d'une téléréunion, inclus dans l'algorithme de la figure 1.

10

La réalisation détaillée d'un système de téléréunion décrite ci-dessous ainsi que ses variantes sont relatives à une téléréunion, dite conférence ou réunion téléphonique organisée entre
15 plusieurs terminaux téléphoniques T_1 à T_I , avec $I \geq 2$. La téléréunion est accessible par réservation ou par abonnement ou immédiatement à travers un pont de conférence téléphonique PCT.

Autour de chaque terminal T_i est installé un
20 groupe G_i de participants $T_{i,1}$ à T_{i,K_i} , où K_i est un entier propre au groupe G_i au moins égal à 1. Par exemple, le terminal T_i peut être une table d'audioconférence autour de laquelle K_i participants prennent place, ou plus simplement un poste
25 téléphonique ou radiotéléphonique.

Ainsi, la configuration minimale d'une téléréunion selon l'invention consiste en deux groupes G_1 et G_2 , avec un participant devant le terminal T_1 et deux participants devant le terminal
30 T_2 .

Le système de téléréunion comprend également un terminal T_0 , confondu avec le terminal T_1 , tel qu'un ordinateur personnel ou une console de téléréunion spécifique, dédié à un organisateur de téléréunion,
35 pour gérer, c'est-à-dire organiser, établir,

surveiller et animer la télé réunion afin notamment d'accueillir les participants, enregistrer leurs identités et des codes d'accès personnels, réserver une télé réunion, filtrer des voies et appeler des participants pendant la télé réunion, organiser des votes, etc. Selon la réalisation illustrée, le terminal TO est en relation avec un serveur local de maintenance et d'assistance SER relié au pont de conférence PCT.

10 Selon une première variante, le pont PCT est relié aux terminaux T_1 à T_I par un réseau numérique à intégration de services RNIS. Par exemple, chaque terminal T_i , avec i compris entre 1 et l'entier I , est un terminal numérique relié à un commutateur du
15 réseau RNIS par une ligne téléphonique lt_i à 2 paires de fils constituant un accès bidirectionnel de base à deux canaux de parole/données B à 64 kbit/s et un canal de signalisation D à 16 kbit/s, ou par un accès bidirectionnel à 30 ou 23 canaux B à 64 kbit/s et un
20 canal de signalisation D à 64 kbit/s lorsque le terminal est par exemple un autocommutateur numérique privé PABX. Un commutateur numérique à autonomie d'acheminement COM du réseau RNIS dessert le pont de conférence PCT par un grand nombre de lignes
25 téléphoniques au moins supérieur à I ; en pratique, un pont peut être desservi par plusieurs centaines de lignes téléphoniques afin d'assurer simultanément plusieurs télé réunions indépendantes. A la figure 1, il est supposé qu'à chaque ligne lt_1 à lt_I correspond
30 une ligne d'accès LT_1 à LT_I au pont PCT, ces lignes pouvant être multiplexées en sortie et démultiplexées en entrée de l'autocommutateur associé COM.

 Selon une deuxième variante, les terminaux T_1 à T_I sont des terminaux analogiques classiquement
35 reliés au réseau téléphonique commuté RTC par des

lignes téléphoniques à deux fils lt_1 à lt_I . De même, le pont de conférence PCT est relié au réseau RTC par au moins I lignes téléphoniques analogiques à deux fils.

5 Selon encore une autre variante, un terminal T_i est un terminal radiotéléphonique portable, par exemple de type GSM, DCS 1800 ou DECT, relié au réseau RTC, RNIS par le réseau fixe de radiotéléphonie correspondant, ou bien une borne
10 privée ou public dans un tel réseau fixe.

 Toutefois, selon d'autres variantes, les variantes précédentes sont combinées. Dans la suite, le pont de conférence PCT est décrit indifféremment du type de ligne téléphonique lt_1 à lt_I et LT_1 à LT_I
15 et du type du réseau RNIS, RTC ; au regard de l'identification d'un intervenant dans une téléréunion sont principalement modifiées des interfaces de réseau IR_1 à IR_I dans le pont de conférence PCT reliées respectivement aux lignes
20 téléphoniques LT_1 à LT_I .

 L'interface de réseau IR_i a pour rôle principal, indépendamment du type de la ligne LT_i , d'extraire de la signalisation téléphonique classique et de la
25 signalisation de téléréunion provenant du terminal T_i dans le signal reçu par la ligne LT_i en les séparant d'un signal de voix de participant, et d'insérer de la signalisation téléphonique et de la signalisation de téléréunion destinée notamment au terminal T_i dans
30 le signal émis dans la ligne LT_i en la mélangeant à un signal de voix.

 Comme montré à la figure 2, un bus de signalisation BSG échange les signalisations extraites et à insérer entre les interfaces IR_1 à IR_I

et une unité de gestion d'appels téléphoniques et de téléréunion UG.

Une voie numérique de réception de parole VR_i relie l'interface de réseau IR_i aux entrées d'un
5 détecteur d'activité de ligne DA et d'un analyseur de voix de groupe AV. Une voie numérique d'émission de parole VE_i relie la sortie d'un mélangeur audio respectif MA_i à l'interface de réseau IR_i .

Lorsque la ligne LT_i est du type numérique,
10 celle-ci supporte au moins l'un des canaux B et le canal D correspondant à ceux de la ligne lt_i . L'interface de réseau IR_i comprend alors notamment un démultiplexeur-multiplexeur pour séparer les voies de parole VR_i et VE_i correspondant au canal B et la
15 signalisation téléphonique et de téléréunion dans le canal D.

Lorsque la ligne LT_i est analogique, l'interface de réseau IR_i comprend notamment plusieurs détecteurs de signalisation téléphonique pour la prise et la
20 libération de ligne et divers signaux de sonnerie, un convertisseur 2 fils/4 fils suivi d'un annuleur d'écho électrique, un modem et/ou un ensemble détecteur/générateur de signaux codés à multifréquence MF pour la signalisation de
25 téléréunion relié au bus BSG, et un codeur MIC et un décodeur MIC reliés respectivement aux voies VR_i et VE_i .

Diverses signalisations particulièrement pour la téléréunion, dite également réunion ou conférence
30 téléphonique, y compris relatives à des commandes d'animation de l'organisateur à travers son terminal TO, et à des demandes/réponses des participants, qui sont susceptibles d'être reçues, traitées et émises par le pont de conférence téléphonique PCT, sont
35 indiquées dans le brevet EP-0 515 241 du demandeur et

dans l'article intitulé "De la Réunion Téléphone à la Conférence Téléphone" par Jean-Pierre BLIN et Jean-Pierre LEBLANC, l'écho des RECHERCHES, N° 171, 2^e trimestre 1998, pages 13 à 20.

5 Outre ces signalisations connues, l'invention introduit un identifiant $ID_{i,k}$ d'un participant-intervenant $P_{i,k}$ appartenant au groupe de participants G_i autour du terminal T_i , dans les signaux émis par les interfaces de réseau IR_1 à IR_I ,
10 y compris donc l'interface de réseau IR_i correspondant à l'intervenant afin qu'il vérifie que l'identifiant transmis aux autres groupes soit correct.

15 Les mélangeurs audio MA_1 à MA_I et l'unité de gestion UG assurent des fonctionnalités connues.

20 Un mélangeur MA_i additionne normalement tous les signaux de parole dans les voies de réception VR_1 à VR_{i-1} et VR_{i+1} à VR_I afin de restituer les signaux audio mélangés représentatifs des voix des participants captées par les terminaux distants du terminal T_i à celui-ci. Le signal de parole reçu dans la voie VR_i est diffusé vers tous les mélangeurs MA_1 à MA_{i-1} et MA_{i+1} à MA_I ; en variante, le signal dans la voie VR_i est également mélangé dans le mélangeur
25 MA_i . En pratique, le mélangeur MA_i traite les signaux à additionner en fonction de certaines caractéristiques de réception du terminal T_i et ne peut additionner que des signaux de voies VR_1 à VR_I sélectionnées en fonction de commandes produites via
30 le bus BSG par l'unité de gestion UG en réponse à des instructions transmises par le terminal d'organisateur TO. Le mélangeur MA_i transmet également des messages vocaux selon l'invention qui
35 sont générés par une messagerie vocale MV sous la commande de l'unité de gestion UG et qui interrompent

momentanément les signaux audio mélangés dans les voies VE_1 à VE_I ou qui viennent en superposition de ceux-ci de manière à entendre à la fois les voix des participants et un message vocal d'identifiant d'intervenant. La messagerie vocale MV transmet des messages préenregistrés aux terminaux et enregistre des réponses codées des participants.

L'unité de gestion UG gère les appels téléphoniques du pont PCT par les terminaux T_1 à T_I , ainsi que les appels de terminaux par le pont PCT sous la commande du terminal TO ou du serveur SER. L'unité UG gère aussi diverses étapes d'une téléconférence, y compris certaines étapes préalables à l'établissement d'une téléconférence, comme la création de modèles vocaux de participants, en fonction de programmes d'instructions établies principalement depuis le terminal d'organisateur TO. En particulier, l'organisateur suit sur l'écran du terminal TO le déroulement d'une téléconférence en affectant à chaque groupe, des identités aux participants, un code d'intervenant, d'auditeur, d'absent, de libération de terminal, etc., en étroite relation avec une mémoire de données de l'unité UG interrogeable depuis le terminal TO ou le serveur SER.

25

Comme cela apparaît à la figure 2, le pont de conférence PCT comprend, en outre selon l'invention, un détecteur d'activité de ligne DA et un analyseur de voix de groupe AV qui sont chacun reliés en écoute aux voies de réception de parole VR_1 à VR_I .

Le détecteur DA détecte parmi les voies de réception, celle qui est la plus active et qui transmet la voix de l'intervenant courant au cours d'une téléconférence. Le détecteur DA établit l'adresse G_i de la voie VR_i ou de la liaison de transmission

35

lt_i - LT_i la plus active, c'est-à-dire du groupe de participants le plus actif. L'adresse G_i constitue une partie de l'identifiant de l'intervenant et est appliquée à l'unité de gestion UG et à l'analyseur AV.

L'analyseur de voix AV sélectionne la voie de réception de parole VR_i correspondant à l'adresse de la plus active liaison G_i établie par le détecteur DA afin d'y extraire un segment vocal SG_i qu'il analyse.

Une mémoire MM enregistre des modèles vocaux des participants à une téléconférence qui sont établis préalablement et/ou au début d'une téléconférence par l'analyseur de voix AV sous la commande de l'unité de gestion UG. Un modèle vocal $MOD_{i,k}$ est associé à l'adresse $P_{i,k}$ du k ème participant appartenant au groupe G_i avec $1 \leq k \leq K_i$.

Au cours d'une période T pendant une téléconférence, l'analyseur AV compare le segment vocal SG_i dans la voie de réception active VR_i avec les modèles de voix $MOD_{i,1}$ à MOD_{i,K_i} , comme cela est décrit ci-après, pour déterminer le participant qui a pris la parole dans le groupe actif G_i afin de l'identifier en tant qu'intervenant au moins par l'adresse $P_{i,k}$, par exemple composée de ses nom et prénom. L'adresse $P_{i,k}$ de l'intervenant est associée à l'adresse G_i de son groupe produite par le détecteur DA afin que l'unité de gestion UG diffuse un identifiant $ID_{i,k} [P_{i,k}, G_i]$ vers tous les autres terminaux T_1 à T_I à travers les interfaces de réseau IR_1 à IR_I .

Comme montré à la figure 3, le procédé d'identification d'intervenant selon une réalisation préférée de l'invention comprend quatre étapes principales E1 à E4 pour créer des modèles vocaux des

participants à une télé-réunion, rechercher une liaison active, c'est-à-dire le groupe auquel un participant, dit intervenant, ayant pris la parole appartient, rechercher l'intervenant dans le groupe
5 actif, et transmettre l'identifiant de l'intervenant par le pont de conférence PCT vers les terminaux T_1 à T_I .

Le modèle vocal d'un participant à une
10 télé-réunion peut être enregistré en différé, antérieurement à la télé-réunion, ou au début de la télé-réunion, comme indiqué à une première sous-étape E10 dans l'algorithme de création de modèle vocal E1 montré à la figure 4.

15 Si l'enregistrement du modèle vocal doit précéder la télé-réunion, le terminal organisateur TO appelle à travers le pont PCT un participant $P_{i,k}$ situé au niveau d'un terminal quelconque connu de l'organisateur, ou bien le participant lui-même $P_{i,k}$
20 appelle depuis un terminal téléphonique quelconque la messagerie vocale MV dans le pont PCT à une sous-étape E11.

A la sous-étape suivante E12, l'organisateur ou la messagerie vocale invite le participant $P_{i,k}$ à
25 prononcer une phrase libre du genre suivant, excédant une durée minimale prédéterminée DP :

"Bonjour, je suis le participant $P_{i,k}$ de l'entreprise $X_{i,k}$ du site G_i ",

dans laquelle $P_{i,k}$ est par exemple les nom et
30 prénom du participant, $X_{i,k}$ le nom de son entreprise et G_i le nom du site du terminal T_i . Un modèle vocal $MOD_{i,k}$ défini par des caractéristiques déterminées de la voix du participant $P_{i,k}$ est construit dans l'analyseur selon un algorithme d'analyse
35 prédéterminée appliqué à la phrase prononcée

précédente. Le modèle vocal $MOD_{i,k}$, dit également référence acoustique, est enregistré avec l'identifiant $ID_{i,k}$ [$P_{i,k}$; $X_{i,k}$; G_i] du participant $P_{i,k}$ dans la mémoire MM à la sous-étape E13.

5 Afin que le modèle vocal $MOD_{i,k}$ puisse servir de référence dans les analyses ultérieures de la voix du participant segmentée périodiquement et particulièrement afin de distinguer cette voix parmi d'autres, la phrase à prononcer est suffisamment
10 longue et dure au moins $DP = 5$ s environ par exemple. Pendant cette durée d'apprentissage DP, le modèle est créé dans l'analyseur AV et répertorié à une adresse correspondant au participant $P_{i,k}$ dans la mémoire MM. La communication téléphonique entre le participant et
15 le pont PCT est ensuite rompue si la télé Réunion est prévue ultérieurement.

 En variante, lorsque l'analyseur AV est incapable de reconnaître l'un des paramètres $P_{i,k}$, $X_{i,k}$, G_i , il le signale au terminal d'organisateur
20 TO ou à la messagerie vocale MV afin d'inviter le participant à répéter la phrase, ou le paramètre incompris, ou afin d'inviter l'organisateur ou le participant à saisir au clavier de son terminal le ou les paramètres incompris $P_{i,k}$, $X_{i,k}$ et G_i en
25 confirmation de la phrase prononcée.

 A la sous-étape E13, si le participant $P_{i,k}$ utilise fréquemment le service de télé Réunion avec le pont PCT, il est déjà répertorié dans la mémoire MM, les paramètres [$P_{i,k}$; $X_{i,k}$; G_i] peuvent être
30 reconnus par l'analyseur de voix AV, et le modèle $MOD_{i,k}$ est mis à jour pour garantir de bonnes performances d'identification vocale au regard d'un changement récent possible de caractéristiques de la voix.

Selon l'alternative à la sous-étape E10, ou après plusieurs ensembles de sous-étapes E11 à E13, l'organisateur initie le début de la télé réunion à la sous-étape E14. En fonction du répertoire des participants de la télé réunion affiché dans le terminal T0, l'organisateur invite l'un après l'autre au moins les participants $P_{i,k}$ dont aucun modèle vocal n'est enregistré ou un modèle vocal ancien est enregistré dans le pont PCT à se présenter aux autres participants, particulièrement aux autres groupes de participants, en prononçant une phrase du genre ci-dessus, et ainsi à effectuer des sous-étapes E15 et E16 analogues aux sous-étapes E12 et E13. La création d'un modèle vocal pour chaque participant est commandée dans l'analyseur AV par l'organisateur depuis le terminal T0 à travers l'unité UG, et la voix du participant se présentant $P_{i,k}$ est diffusée vers les terminaux T_1 à T_I par des commandes appropriées de l'unité UG depuis le terminal T0.

La création de modèles vocaux aux sous-étapes E15 et E16 juste au début de la télé réunion améliore les performances d'identification vocale par l'analyseur AV au cours de la télé réunion, comparativement à une création préalable des modèles vocaux aux sous-étapes E12 et E13 pour une même durée de parole servant à la création du modèle. Le laps de temps entre la phase d'apprentissage E15-E16 et l'étape d'identification E3 proprement dite, c'est-à-dire de reconnaissance de voix, est très réduit, ce qui amoindrit notablement les effets perturbants dûs à des changements, évolutions et vieillissements de la voix du participant. La phase d'apprentissage E15-E16 et l'étape d'identification E3 proprement dite se déroulent à travers le même terminal T_i et la même liaison de transmission lt_i-LT_i entre le terminal T_i

et le pont PCT, ce qui maintient les mêmes prise de son et mode de transmission téléphonique au cours des identifications du participant, en tant qu'intervenant pendant la télé Réunion.

5 Au début de la télé Réunion, lorsque tous les participants sont recensés et leurs modèles vocaux de référence sont établis et enregistrés à la dernière sous-étape E17, les participants peuvent converser librement, les intervenants étant identifiés au fur
10 et à mesure du déroulement de la télé Réunion aux étapes cycliques E2, E3 et E4 décrites ci-après, comprenant respectivement des sous-étapes E20 à E22, E30 à E36, et E40 et E41.

15 Dans la suite, en référence à la figure 5, il est supposé que le participant $P_{i,k}$ dans le groupe G_i à proximité du terminal T_i prend la parole. On remarquera, en pratique, que dans l'algorithme de la figure 5 principalement mis en œuvre dans le
20 détecteur d'activité DA et l'analyseur de voix AV, le résultat de chacune des sous-étapes de détermination de puissance moyenne ou de distance dans les étapes E2 et E3 est issu d'une convergence de plusieurs sous-résultats selon des critères prédéterminés
25 pendant une période variable de 0,1 s à 5 s environ, chaque sous-résultat étant évalué pendant une sous-période de 10 ms à 20 ms environ, afin de déterminer pendant chaque période, le groupe actif et l'intervenant dans celui-ci. La période T est
30 suffisante pour analyser des segments vocaux dans les voies de réception VR_1 à VR_I et y reconnaître une voix particulière d'intervenant.

 L'algorithme de la figure 5 contient quelques sous-étapes incluant une détermination de distance ;
35 plus généralement, cette détermination peut être

remplacée par une mesure de similarité ou de vraisemblance entre un segment vocal et un modèle vocal.

5 Préalablement, en début de réunion, un modèle de dernier intervenant MOD(DI) et mis à zéro.

La recherche d'une ligne active repose sur un découpage des signaux vocaux dans les voies de réception VR_1 à VR_I en des segments vocaux parallèles SG_1 à SG_I pendant chaque période T, à la sous-étape
10 E20. Ce découpage est réalisé dans le détecteur DA et l'analyseur AV.

Le détecteur d'activité DA moyenne les puissances PM_1 à PM_I des segments SG_1 à SG_I pendant
15 la période T et détermine la puissance moyenne maximale pendant cette période, sous réserve qu'elle soit supérieure à un premier seuil SE1, à la sous-étape E21. Si la puissance moyenne maximale est inférieure à SE1, le détecteur d'activité DA confirme
20 un "silence" au cours de la télé-réunion à l'unité de gestion UG qui ne diffuse aucun identifiant. L'algorithme retourne normalement à la sous-étape E20, à travers une sous-étape finale EF si la télé-réunion n'est pas terminée, ou si le procédé
25 d'identification n'est pas interrompu dans le pont PCT par l'organisateur notamment. Lorsque le "silence" se prolonge ou un faible bruit persiste, l'organisateur peut intervenir pour inviter les participants à parler plus fort, en complément aux
30 commandes automatiques de gain incluses aux extrémités des liaisons $lt_1 - LT_1$ à $lt_I - LT_I$ entre pont et terminaux.

A la sous-étape E22, le détecteur AV identifie la voie de réception VR_i et donc la ligne active LT_i
35 qui correspondent à la puissance moyenne maximale

précédemment déterminée. Le détecteur AV fournit alors le nom du groupe actif G_i associé à la ligne active LT_i en tant que partie de l'identifiant de intervenant à l'analyseur AV. L'analyseur et l'unité
5 de gestion UG adressent en lecture l'ensemble des modèles vocaux $MOD_{i,1}$ à MOD_{i,K_i} des K_i participants du groupe G_i dans la mémoire MM et sélectionne le segment SG_i dans la voie de réception VR_i pour l'analyser à l'étape E3.

10 L'analyseur de voix AV recherche alors l'intervenant $P_{i,k}$ dans le groupe actif G_i selon les sous-étapes E30 à E36 de l'étape principale E3 dont les sous-étapes E30 à E33 sont optionnelles.

15 De préférence préalablement, l'analyseur vérifie à une sous-étape E30 qu'il est en mesure de distinguer une voix de participant dans le segment vocal SG_i . Cette vérification consiste à déterminer la distance entre des caractéristiques prédéterminées
20 du segment SG_i , exprimées sous forme de composantes vectorielles, et celles d'un modèle générique représentatif d'une prise de parole multiple $MOD(PM, K_i)$ pour le nombre de participants K_i dans le groupe G_i . Si la distance $\|SG_i, MOD(PM, K_i)\|$ est
25 inférieure à un seuil prédéterminé $SE2$, le segment SG_i contient plusieurs voix de participants du groupe G_i ayant pris la parole simultanément pendant la période T. Ceci est significatif d'un brouhaha dans lequel l'analyseur AV est incapable de sélectionner
30 un locuteur prépondérant. A la sous-étape E31, l'analyseur AV commande alors l'unité UG pour que la messagerie vocale MV transmette un message vocal d'alerte du type : "Plusieurs personnes parlent dans le groupe G_i " ou "Ne parlez pas tous en même temps
35 dans le groupe G_i ", ou pour que l'unité UG transmette

un message affichable analogue au message vocal, au moins dans la voie d'émission VE_i . L'algorithme va ensuite à l'étape EF.

Lorsque la distance déterminée à la sous-étape
 5 E30 est supérieure au seuil SE2, l'analyseur AV vérifie que l'intervenant courant ayant émis le segment SG_i est le même que pendant la période précédente T à la sous-étape E32. L'analyseur compare la distance entre le segment courant SG_i et le modèle
 10 MOD(DI) du dernier intervenant DI qui peut appartenir à n'importe quel groupe, y compris le groupe G_i , à un seuil prédéterminé SE3 qui peut être égal au seuil SE2, en sachant qu'au début de la télé-réunion, la distance entre le segment SG_i du premier intervenant
 15 et MOD(DI) = 0 est considérée comme très supérieure à SE2. A l'étape E32, selon une autre variante, le segment courant SG_i et le segment SG_i à la période précédente T sont comparés pour y détecter une rupture de signal vocal indiquant un changement de
 20 locuteur ; ou bien ces deux variantes sont combinées.

Si cette distance est inférieure au seuil SE3, l'intervenant n'a pas changé à la sous-étape E33. L'unité de gestion UG n'a pas besoin de transmettre à nouveau l'identifiant du dernier intervenant.
 25 Toutefois, l'unité UG rafraîchit la mémoire des participants en diffusant périodiquement l'identifiant du dernier intervenant vers les terminaux T_1 à T_I , par exemple toutes les $P = 5$ secondes environ, tant que ledit dernier intervenant
 30 parle.

Selon une variante de la sous-étape E33, l'unité UG retransmet l'identifiant du dernier intervenant qu'aux participants dont les terminaux ont émis une demande de transmission d'identifiant d'intervenant
 35 au moyen d'un code spécifique composé sur les

claviers de ces terminaux. Cette variante permet à un participant distrait ou momentanément absent de connaître l'identité du dernier intervenant, notamment lorsque l'intervention de celui-ci est
5 particulièrement longue.

Lorsque le dernier intervenant DI a cessé de parler à la sous-étape E32, l'analyseur AV détermine à la sous-étape suivante E34 les distances entre le segment SG_i et les modèles $MOD_{i,1}$ à MOD_{i,K_i} des K_i
10 participants dans le groupe actif G_i désigné par le détecteur DA à la sous-étape E22. L'analyseur AV compare la plus petite de ces distances avec un seuil prédéterminé SE4 qui peut être égal au seuil SE3 ou SE2. Si la plus petite distance $\|SG_i, MOD_{i,k}\|$ est
15 supérieure à SE4 avec $k \in [1, K_i]$, l'analyseur AV commande à l'unité de gestion UG de faire diffuser par la messagerie MV un message vocal, ou par l'interface IR_i un message affichable "intervenant inconnu" à la sous-étape E35, l'algorithme passant
20 ensuite à l'étape EF.

Par contre, lorsque la plus petite distance précitée est inférieure à SE4, c'est-à-dire lorsque le segment SG_i présente une certaine similarité ou cohérence avec le modèle $MOD_{i,k}$, l'analyseur lit dans
25 la mémoire MM la désignation $P_{i,k}$ du participant correspondant, éventuellement accompagnée d'autres paramètres comme le nom de l'entreprise $X_{i,k}$ à la sous-étape E36, pour les appliquer à l'unité de gestion UG qui les associe au paramètre G_i pour
30 constituer l'identifiant $ID_{i,k}$.

Au cours de la sous-étape E36, l'unité de gestion UG ayant reçu les paramètres G_i du détecteur DV et $P_{i,k}$ et $X_{i,k}$ de l'analyseur AV compose un message vocal ou visuel contenant l'identifiant $ID_{i,k}$
35 $[P_{i,k}, X_{i,k}, G_i]$.

En général, ce message est diffusé vers tous les terminaux T_1 à T_I à travers les interfaces de réseau correspondants à la sous-étape E40. Ce message, comme
5 les autres messages de télé Réunion établis par l'unité de gestion UG aux sous-étapes E31, E33 et E35, est transmis :

- soit sous forme de signal numérique en coupure des signaux vocaux dans les voies d'émission
10 correspondantes VE_1 à VE_I afin d'être restitué à la sous-étape E41 par des écouteurs ou des haut-parleurs individuels et/ou des haut-parleurs communs dans les terminaux,

- soit sous forme de signal visuel par
15 multiplexage numérique ou fréquentiel en fonction du mode de transmission du réseau RNIS ou RTC afin d'être visualisé à la sous-étape E41 dans les terminaux, par exemple dans des afficheurs de postes téléphoniques ou radiotéléphoniques ou dans des
20 écrans de terminaux ; les messages visuels sont exprimés en caractères alphanumériques avec ou sans symboles ; par exemple, un point d'interrogation signifie "Intervenant inconnu" à la sous-étape E35.

Ces messages sont également transmis par le pont
25 PCT au terminal $TO = T_1$ pour que l'organisateur puisse suivre et animer la télé Réunion ; les identifiants garantissent l'identité des participants, par exemple à l'occasion d'un vote, ce qui exclut toute falsification d'identité en cours de
30 télé Réunion. Ces messages sont également transmis au serveur de conférence SER pour établir des statistiques notamment sur les temps de parole de chaque participant, ou de chaque groupe.

Si le réseau de transmission est de type RNIS,
35 les messages de télé Réunion affichables, tels que

ceux contenant un identifiant, établis par l'unité UG sont de préférence compris dans de la signalisation d'usager à usager transférée à travers le canal D. Les messages affichables sont ainsi inclus dans des messages d'information d'usager échangés entre des usagers, en l'espèce le pont de conférence PCT et un terminal T_1 , et transportant jusqu'à 128 octets utiles.

Lorsque le réseau de transmission est de type analogique RTC, les messages de télé Réunion affichables, tels que ceux contenant un identifiant, sont analogues au paramètre du nom du demandeur (Calling Party Name), référencé 07h selon la norme ETSI ETS 300 659-1 pour identifier le demandeur lors d'une présentation d'appel au terminal demandé. Selon l'invention, les messages affichables sont adressés par le pont de conférence aux terminaux sous la forme de paramètres non encore utilisés par les normes, tels que les paramètres F1h à FFh selon la norme ETSI précitée. Ces paramètres en mode décroché sont transmis au moyen de modems inclus dans les interfaces de réseau IR_1 à IR_I , par exemple en modulation de fréquence à cohérence de phase selon la recommandation V23 de l'UIT-T.

En variante, le message vocal ou visuel contenant l'identifiant d'intervenant $ID_{i,k}$ n'est diffusé que vers des terminaux prédéterminés à leurs demandes à la sous-étape E40. Chaque terminal qui souhaite connaître l'identité de l'intervenant transmet un message de demande d'identité d'intervenant, par exemple à la suite d'un code spécifique composé au clavier, comme déjà indiqué en variante pour la sous-étape E33. Les messages de demande sont écrits en mémoire de l'unité de gestion UG en dépendance des adresses des terminaux

correspondants. L'unité UG lit les messages de demande à chaque période T pour n'appliquer le message vocal ou visuel d'identifiant d'intervenant qu'aux ensembles d'interface-mélangeur IR_1-MA_1 à IR_I-MA_I correspondant aux terminaux correspondants.

Bien que l'invention ait été décrite ci-dessus relativement à un pont de conférence très éloigné des terminaux, celle-ci n'est pas limitée à cette réalisation mais englobe d'autres répartitions des terminaux et des fonctionnalités incluses dans le pont de conférence. Par exemple, le pont de conférence est un périphérique d'un commutateur téléphonique privé PABX desservant les terminaux. Selon un autre exemple, les fonctionnalités du pont de conférence sont incluses dans le terminal d'organisateur TO, c'est-à-dire dans l'un $T_1 = TO$ des terminaux. En dépendance de liaisons, y compris de type radiotéléphonique, à l'entité incluant les fonctionnalités du pont de conférence selon l'invention, les interfaces de réseau IR_1 à IR_I peuvent être réunies en une seule interface reliée à une seule liaison de transmission analogique ou numérique dans laquelle les voies VR_1 à VR_I et VE_1 à VE_I sont multiplexées fréquemment ou numériquement.

Dans le cas de la configuration minimale avec un groupe G_1 avec un participant $P_{1,1}$ et un groupe G_2 avec deux participants $P_{2,1}$ et $P_{2,2}$, les fonctionnalités du pont de conférence selon l'invention sont de préférence incluses dans le terminal T_1 , ce qui permet au participant $P_{1,1}$ de distinguer la voix du participant $P_{2,1}$, tel que le propriétaire du terminal T_2 , par exemple un

2799914

radiotéléphone, d'un utilisateur occasionnel $P_{2,2}$ du terminal T_2 .

REVENDECATIONS

1 - Procédé pour identifier un intervenant parmi des participants répartis en des groupes (G_1 à G_I) à proximité de terminaux téléphoniques (T_1 à T_I) lors d'une téléréunion gérée par l'intermédiaire d'un dispositif de conférence (PCT) relié aux terminaux à travers des liaisons de transmission ($lt_1 - LT_1$ à $lt_I - LT_I$), caractérisé en ce qu'il comprend, après identification (E3) de la voix de l'intervenant ($P_{i,k}$) en réception dans les liaisons par le dispositif de conférence (PCT), les étapes de :

diffuser (E40) un identifiant ($ID_{i,k}$) de l'intervenant identifié dans les liaisons depuis le dispositif de conférence (PCT) vers les terminaux (T_1 à T_I), et

restituer l'identifiant ($ID_{i,k}$) de l'intervenant dans les terminaux (T_1 à T_I).

2 - Procédé conforme à la revendication 1, selon lequel l'identifiant ($ID_{i,k}$) de l'intervenant comprend au moins le nom du participant ($P_{i,k}$), de préférence complété par au moins une désignation du groupe (G_i) auquel il appartient.

3 - Procédé conforme à la revendication 1 ou 2, selon lequel l'identifiant ($ID_{i,k}$) de l'intervenant est obtenu par reconnaissance de parole dans le dispositif de conférence (PCT).

4 - Procédé conforme à l'une quelconque des revendications 1 à 3, selon lequel l'identifiant ($ID_{i,k}$) n'est diffusé que vers des terminaux en réponse respectivement à des demandes de ceux-ci au dispositif de conférence.

5 - Procédé conforme à l'une quelconque des revendications 1 à 4, selon lequel l'identifiant ($ID_{i,k}$) de l'intervenant est diffusé sous forme de message vocal en coupure ou en superposition de signaux vocaux de participant dans les liaisons ($lt_1 - LT_1$ à $lt_I - LT_I$) et est restitué acoustiquement par les terminaux.

6 - Procédé conforme à l'une quelconque des revendications 1 à 4, selon lequel l'identifiant ($ID_{i,k}$) de l'intervenant est diffusé sous forme de message de signalisation dans les liaisons et est restitué visuellement dans les terminaux.

7 - Procédé conforme à l'une quelconque des revendications 1 à 6, selon lequel le dispositif de conférence (PCT) est inclus dans l'un des terminaux.

8 - Procédé conforme à l'une quelconque des revendications 1 à 6, selon lequel le dispositif de conférence (PCT) est inclus dans un commutateur téléphonique privé desservant les terminaux.

9 - Procédé conforme à l'une quelconque des revendications 1 à 8, caractérisé en ce que l'identification de l'intervenant dans le dispositif de conférence (PCT) comprend les étapes de :

- rechercher (E21) en réception la plus active des liaisons ($lt_1 - LT_1$ à $lt_I - LT_I$),
- identifier (E22) le groupe (G_i) correspondant à la plus active liaison en réception (VR_i),
- déterminer (E34) la plus grande similitude entre un segment vocal (SG_i) dans la plus active

liaison et des modèles vocaux ($MOD_{i,1}$ à MOD_{i,K_i}) des participants appartenant au groupe identifié (G_i), et

5 - identifier (E36) le participant ($P_{i,k}$) correspondant à la plus grande similitude déterminée, en tant qu'intervenant.

10 10 - Procédé conforme à la revendication 9, selon lequel l'étape de rechercher en réception la plus active des liaisons (E21) consiste à rechercher la plus grande des puissances moyennes de segments vocaux (SG_1 à SG_I) dans les voies de réception (VR_1 à VR_I) des lignes de transmission à condition qu'elle soit supérieure à un seuil prédéterminé (SE1).

15 11 - Procédé conforme à la revendication 9 ou 10, selon lequel le participant ($P_{i,k}$) correspondant à la plus grande similitude déterminée n'est pas identifié et un message d'intervenant inconnu (E35) est diffusé vers les terminaux lorsque la plus grande
20 similitude déterminée est inférieure à un seuil prédéterminé.

25 12 - Procédé conforme à l'une quelconque des revendications 9 à 11, selon lequel les modèles vocaux ($MOD_{1,1}$ à MOD_{I,K_I}) des participants sont créés et mis à jour (E1) dans le dispositif de conférence au plus tard au début de la télé-réunion (E11-E13), et résultent de l'analyse vocale d'un genre de phrase excédant une durée minimale prédéterminée, de
30 préférence contenant au moins une partie d'identifiant de participant ($P_{i,k}$).

35 13 - Procédé conforme à l'une quelconque des revendications 9 à 12, comprenant une étape (E30) de déterminer la similitude entre le segment vocal (SG_i)

dans la plus active liaison en réception (VR_i) et un modèle vocal ($MOD(PM, K_i)$) de prise de parole multiple, de préférence spécifique au nombre de participants dans le groupe identifié (G_i), afin que
5 le dispositif de conférence (PCT) diffuse (E31) un message de prise de parole multiple vers les terminaux (T_1 à T_I) lorsque ladite similitude déterminée est supérieure à un seuil prédéterminé.

10 14 - Procédé conforme à l'une quelconque des revendications 9 à 13, comprenant une étape (E32) de déterminer la similitude entre le segment vocal (SG_i) dans la plus active liaison en réception (VR_i) et le modèle vocal ($MOD(DI)$) d'un dernier intervenant afin
15 que le dispositif de conférence (PCT) ne diffuse aucun identifiant, ou ne diffuse (E33) l'identifiant ($ID(DI)$) du dernier intervenant que périodiquement vers les terminaux ou que vers des terminaux respectivement en réponse à des demandes de ceux-ci
20 au dispositif de conférence (PCT), lorsque ladite similitude est supérieure à un seuil prédéterminé.

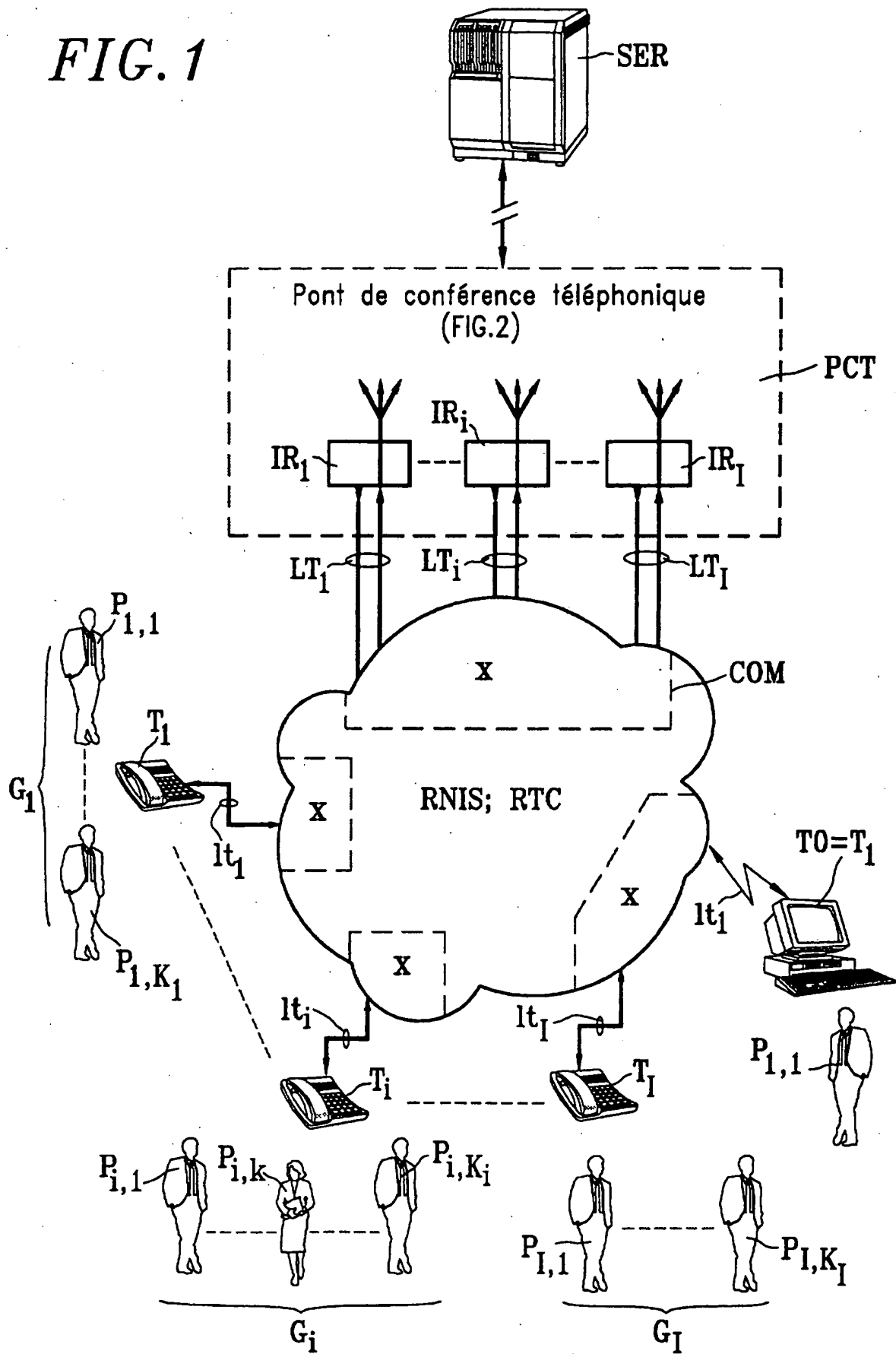
15 - Dispositif de conférence pour la mise en œuvre du procédé conforme à l'une quelconque des
25 revendications 1 à 14, comprenant un moyen (DA) pour détecter périodiquement la plus active liaison en réception (VR_i) afin d'identifier le groupe (G_i) correspondant à la plus active liaison en réception, un moyen (AV) pour comparer périodiquement un segment
30 vocal dans la plus active liaison en réception avec des modèles vocaux mémorisés ($MOD_{i,1}$ à MOD_{i,K_i}) des participants appartenant au groupe identifié (G_i) afin d'identifier le participant ($P_{i,k}$), en tant qu'intervenant, correspondant à la plus grande
35 similitude entre ledit segment vocal et les modèles

vocaux, un moyen (UG) pour établir un identifiant ($ID_{i,k}$) de l'intervenant identifié, et un moyen (IR_1 à IR_I) pour diffuser l'identifiant d'intervenant dans des liaisons ($lt_1 - LT_1$ à $lt_I - LT_I$) vers des terminaux (T_1 à T_I).

5

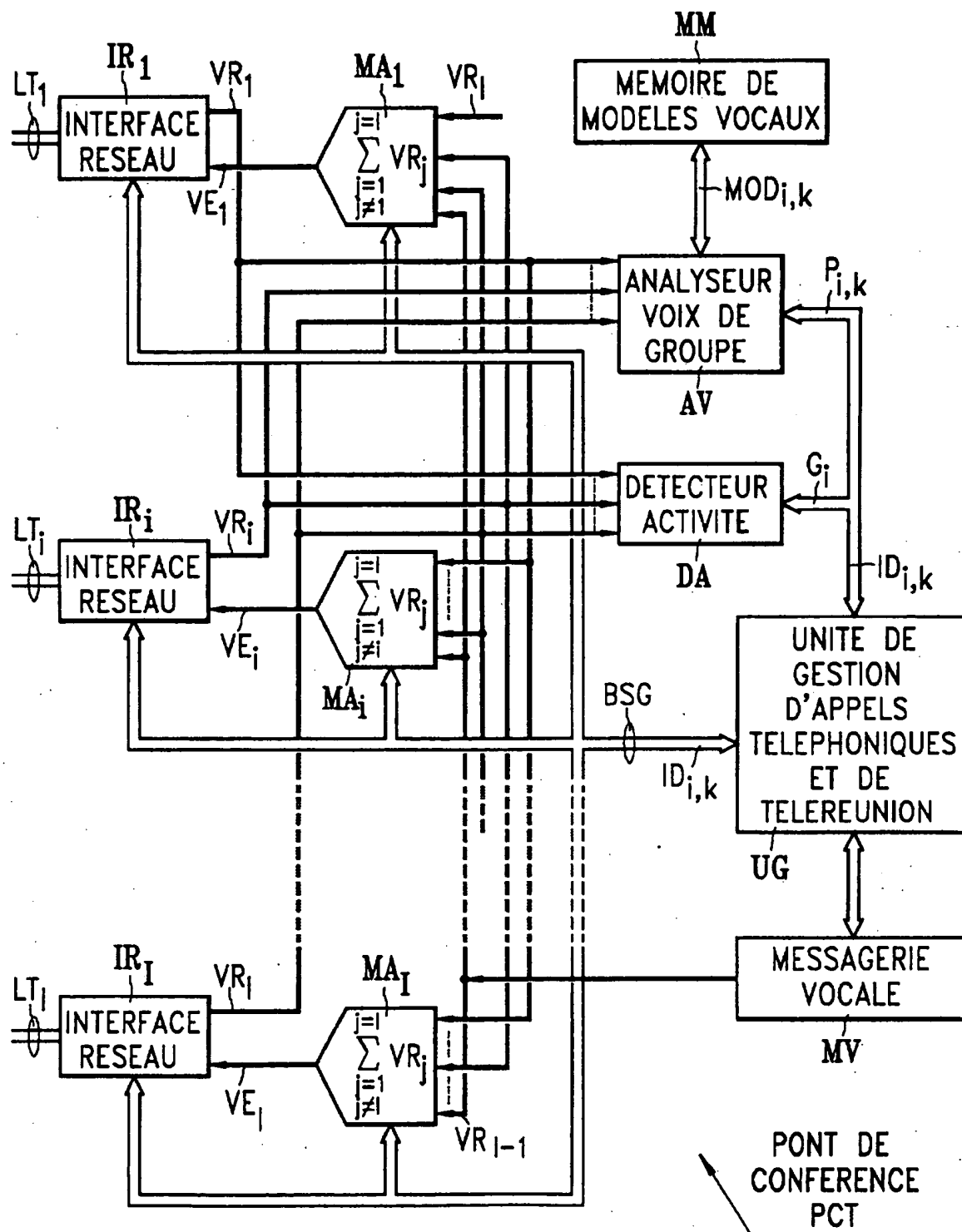
1/5

FIG. 1



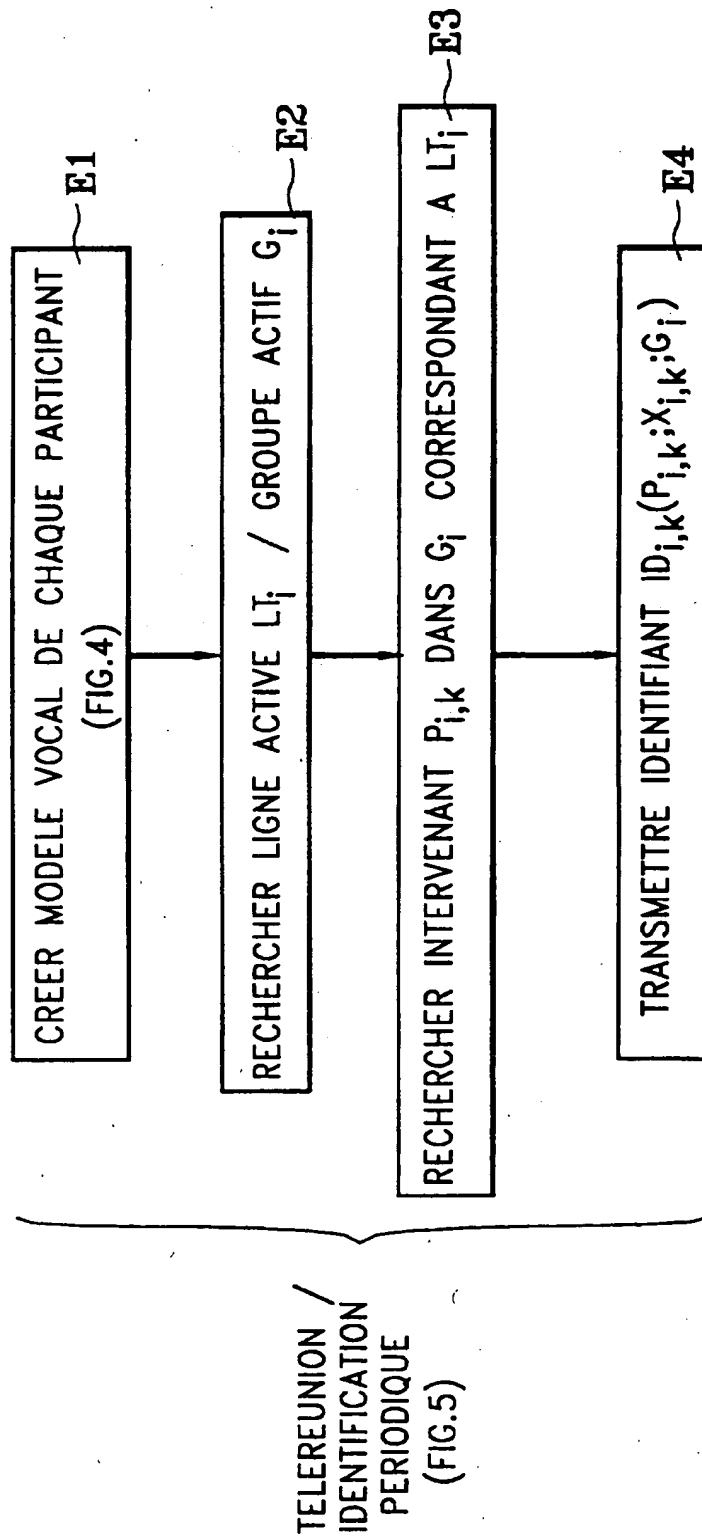
2/5

FIG. 2



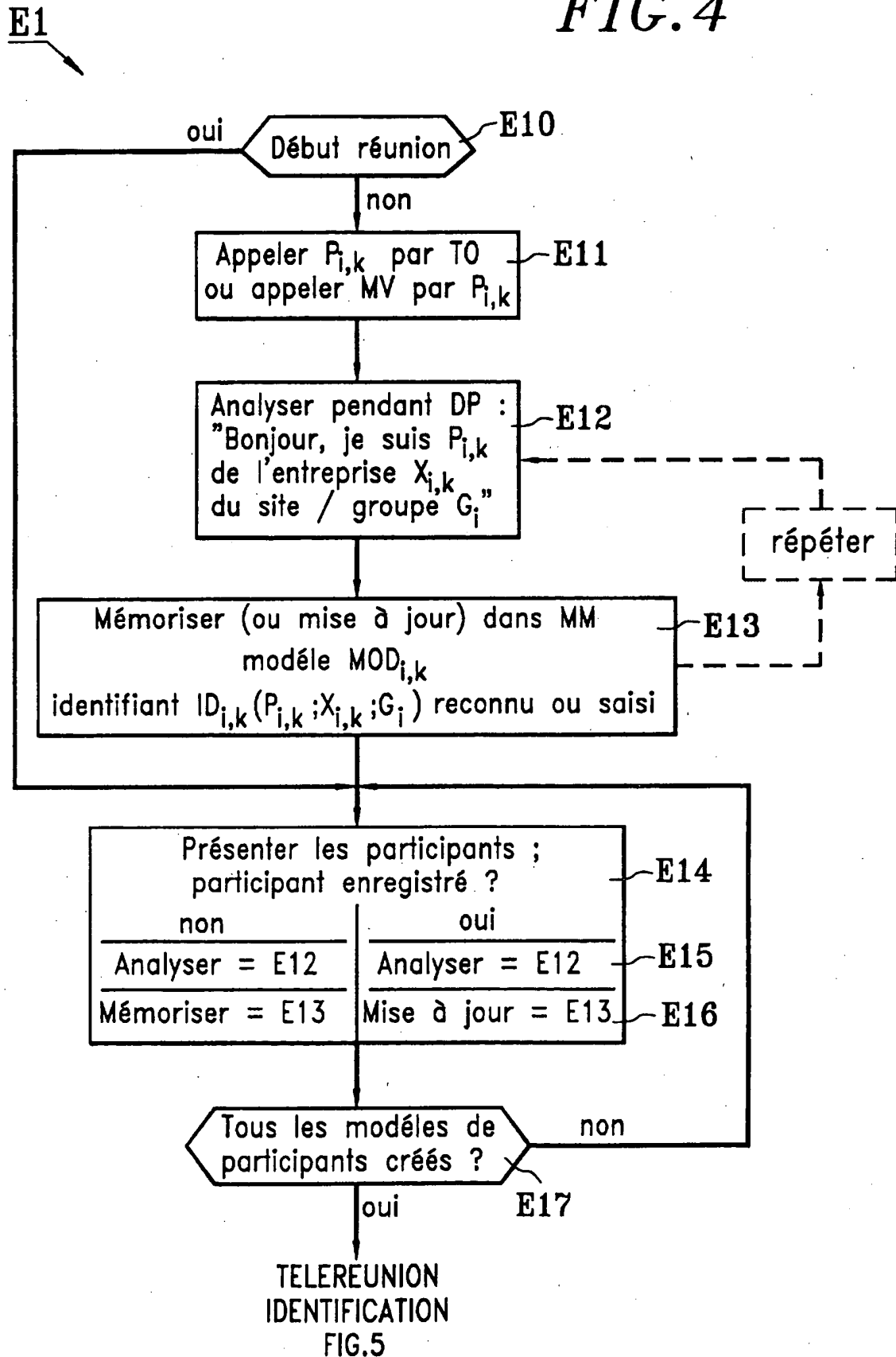
3/5

FIG. 3



4/5

FIG. 4



5/5

CREATION
MODELES VOCAUX
FIG.4

FIG.5

